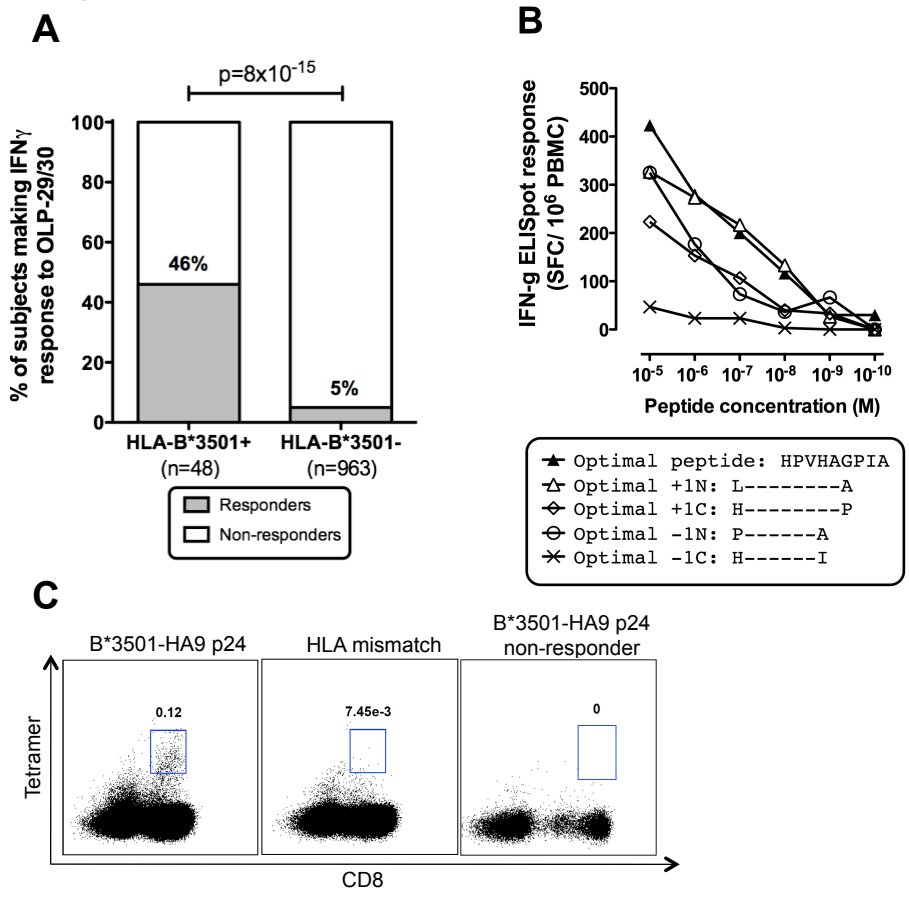
Suppl. Fig 1



Suppl. Table 1

Location of cohort	Clade	Number of subjects	Number (%) with HLA-B*3501	Viral load (copies /ml plasma) Median	Viral load (copies /ml plasma) IQR	CD4+ T cell count (cells/ml) Median	CD4+ T cell count (cells/ml) IQR
Kumamoto, Japan	B	242	37 (15.3)	19500	2575-92000	278	127-469
Mexico City, Mexico	B	771	72 (9.3)	40013	10199-124000	399	157-593
Durban, South Africa	C	1218	46 (3.8)	38200	7315-154250	376	239-519
Gaborone, Botswana	C	514	26 (5.1)	19100	3920-78200	342	220-476

Suppl. Table 2

Protein	OLP number	Clade	OLP sequence	p[a]	q[a]	Optimal epitope P2 ↓	Optimal epitope C-terminus ↓	Epitope name
p24 Gag	29	B	A A E W D R L H P V H A G P I A	8.91E-13	1.73E-19	H **P** V H A G P I	**A**	Gag-HA9[b]
		C	- - - - - - - - - - - - - - - -			- - - - - - - -	-	
p24 Gag	30	B	L H P V H A G P I A P G Q M R E P R	5.67E-06	9.46E-05	H **P** V H A G P I	**A**	
		C	- - - - - - - - - - - - - - - - - -			- - - - - - - -	-	
Nef	85	B	R Y P L T F G W C F K L V P V	5.59E-07	7.28E-07	Y **P** L T F G W C	**Y**	Nef-YY9
		C	- - - - - - - - - - - - - - - -			- - - - - - - -	**F**	Nef-YF9[b]
Rev	95	B	D E E L L K T V R L I K F L Y	1.55E-07	3.50E-08	K **T** V R L I K F L	**Y**	Rev-KY10
		C	- - A - - Q A - - I - - I - -			Q **A** - - I - - I -	-	Rev-QY10[b]
RT	223	B	Q L E K E P I V G A E T F Y V D G A	9.72E-08	1.32E-08	E **P** I V G A E T F	**Y**	RT-EY10[b]
		C	- - - - - - - A - - - - - - - - -			- - - A - - - - -	-	
Int	252	B	G Y I E A E V I P A E T G Q E T A Y	1.41E-09	1.07E-12	I **P** A E T G Q E T A **Y**		Int-IY11
		C	- - - - - - - - - - - - - - - - - - -			- - - - - - - - - -	-	

[a] p and q values for association between expression of HLA-B*3501 and response to OLP computed from analysis of ELISpot data from 1009 subjects
[b] Location of epitope also defined by identification of sequence polymorphism associated with HLA-B*3501 in 710 C-clade infected subjects in Durban, South Africa

Suppl. Table 3

Redidue at position Gag260 (HXB2)	HXB2 Position	Target amino acid	Consensus amino acid	Direction of the association with Gag260	N	P-value	Q-value	Conditions[a]
E	146	P	A	Negative	1817	1.2E-06	8.7E-04	B*57, Cw*06, 242N, 168I, 149A
E	168	I	V	Negative	1840	1.1E-10	1.8E-07	
D	207	D	E	Negative	1852	3.1E-06	2.0E-03	
D	215	V	L	Negative	1853	7.0E-12	1.3E-08	
E	215	I	L	Positive	1853	3.8E-07	3.1E-04	256I
D	228	V	M	Negative	1853	3.3E-07	2.7E-04	
D	250	I	M	Negative	1856	5.1E-15	1.5E-11	
E	256	V	I	Positive	1823	7.1E-07	5.4E-04	250I, 138A
E	268	M	L	Negative	1856	5.9E-06	3.4E-03	

[a] These variables were added to the model before residue at Gag260 was added

Suppl. Figure 1: Unequivocal confirmation of HA9 (HPVHAGPIA, Gag-216-224) as an optimal epitope restricted by HLA-B*3501.

A: Significant associations between IFNγ ELISpot responses to either OLP-29 and/or OLP-30 (containing the HA9 peptide) and expression of HLA-B*3501; pooled data from subjects in Durban, South Africa (C-clade infected, n=795) and the Thames Valley Cohort (mixed clades, n=214). Of these, 48 (4.7%) had HLA-B*3501. p value by Fisher's exact test.

B: IFNγ ELISpot responses made by an HLA-B*3501-positive adult subject with chronic B-clade HIV-1 infection (Thames Valley subject R051, HLA-A*0101, -A*3002, -B*1801, -B*3501, -Cw*0401, -Cw*0501) in response to the optimal epitope HA9 and four peptide truncations of this peptide.

C: HLA-B*3501-HPVHAGPIA (HA9) tetramer staining of B-clade HIV-1 infected subject (Thames Valley subject N030, HLA-A*0201, –A*0301, -B*3502, -B*4402, -Cw*0401, -Cw*0501) responding to OLP-29/30 (left) controlled by HLA-mismatch tetramer (HLA-B*4402) (middle) and a non-HA9 responder subject (Thames Valley subject OX035, HLA-A*0201, –A*1101, -B*1801, -B*3501, -Cw*0401, -Cw*0501) stained with HLA-B*3501-HA9 tetramer (right).

Suppl. Table 1: Characteristics of unpublished study cohorts

Suppl. Table 2: Novel HLA-B*3501 restricted optimal epitopes defined by IFNγ ELISpot screening and sequence analysis, combined with motif inference.

p/q values for HLA-B*3501 association with the optimal determined from 1009 adult subjects with chronic HIV-1 infection. Sites of polymorphisms are previously published, identified by lineage-corrected sequence analysis of Durban cohort (7).

Suppl. Table 3: Covariation at residue Gag-260

Covariation analysis (phylogenetic dependency network) for 9 high-variability p24 residues in >1,800 sequences from C-clade infected individuals from Botswana, Zimbabwe and Durban, South Africa.